Applicability of Network Logs for Securing Computer Systems

Nikita Singh[1], Pavitra Chauhan[2], and Nidhi Chandra[2]
[1] Amity University, Computer Science Department, Noida, India
Email: nikita.sngh1@gmail.com
[2] Amity University, Computer Science Department, Noida, India
Email: {chauhan.pavitra20, srivastavanidhi8}@gmail.com

***Abstract*—Logging the events occurring on the network has become very essential and thus playing a major role in monitoring the events in order to keep check over them so that they doesn't harm any resources of the system or the system itself. The analysis of network logs are becoming the beneficial security research oriented field which will be desired in the computer era. Organizations are reluctant to expose their logs due to risk of attackers stealing the sensitive information from their respective logs. In this paper we are defining architecture and the security measures that can be applied for a particular network log.**

***Index Terms*—Log based systems, Network Logs, Log management**

I. Introduction

For securing a Computer System, a well-defined security policy is needed. A security policy can be defined as the framework within which an organization establishes required levels for securing the systems to achieve the desired confidentiality targets. A policy is a statement of information values, protection responsibilities and organization commitment for a system. A security policy varies for each organisation; a common policy standard cannot be defined. A security policy for one organization may not be sufficient for another. A well-defined and standard security policy can also be formulated using logs. Log data plays a very important role for security researchers, organizations to enrich network management and security standards. A proper assessment of logs is very crucial for understanding the alerts caused by the intrusions.

A log keeps an account of all the events and actions that occurs on a system. Even if an activity or an event goes unnoticed on a system, logs helps to trace them for identifying various intrusions as close to real-time as possible. There are various sources of logs in a host system and in a network, and are in different formats. Some of the log source includes:

- *Logs from the Host System:* There are three types of logs that are generated in the event viewer of the windows by default:
  i. System logs: includes the events logged by the system components e.g. functioning of drivers.
  ii. Application logs: includes events logged by the application or programs depending on the developer of the software program e.g. database program may record file error.
  iii. Security logs: includes events such as valid or invalid logon or events related to creating, deleting or opening of files.
- *Logs from the network:* There are several ways of obtaining the logs i.e. recording of the events occurred in the network. Some of the logs that can be obtained are:
  i. Firewall logs
  ii. IDS/IPS logs
  iii. Application Server logs

The computer and network logs are being generated today at a faster pace and they are being analysed using several tools developed by various companies. Security log analysis systems are also known as log-based Intrusion Detection systems (LIDS). Such systems are becoming the need of today's computer users because they are the one whose analysis keeps the system secure. Log Analysis for Intrusion Detection is the process or techniques used to detect attacks on a specific environment using logs as the primary source of information and they are also used to detect computer misuse, policy violation and other forms of inappropriate activities [1]. Dr. Kees in [1] have discussed about organizational view where he states that one or more centralized logging servers and configure logging devices throughout organizations to send duplicate of their log entries to the centralized logging servers. For performing all such activities we require log management infrastructure which comprises of the hardware, software, networks and media used to generate, transmit, store, analyse and dispose of log data.

In [2] the author has shown recent scenario of today's networking environment which can be visualised as follows in figure 1.

It has been concluded that the security appliances need constant signature updates on attacks in order to work properly for unknown attacks. As the usage of computer system is increasing, the data being generated using logs are also increasing and large data presents bigger risk. To reduce the risk factor of the data being leaked, analysis of such data at frequent interval is necessary.

The whole paper is divided into several sections where Section II is the background about logs and their types are also discussed, Section III covers the related work based on logs, on other hand Section IV consist of how log management came into picture, Section V gives the basic architecture of

DOI: 01.IJNS.4.1.18

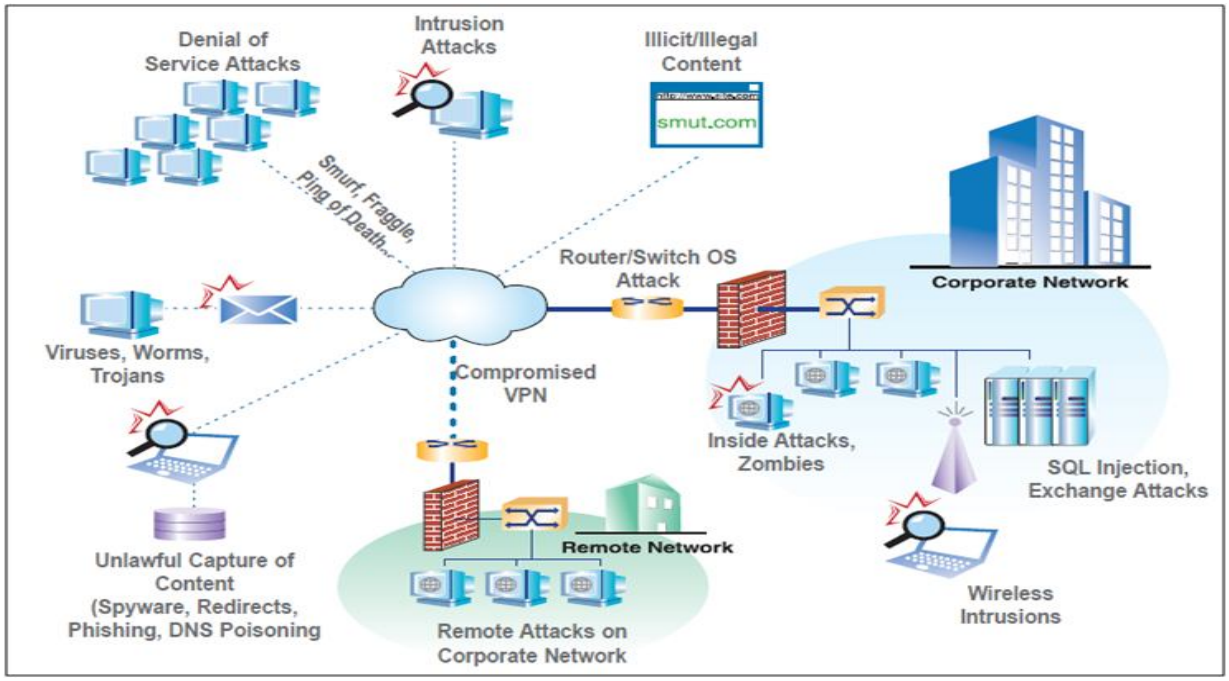

Figure 1. Prevalent threats vectors in today's Networking Environment

both host as well as network logs and in section VI gives the future scope of the topic.

II. Background

A network log contains all the information about a network that is being analysed. This information is formatted on the basis of various header fields such as Source IP address, Destination IP address, Source port number, Destination port number, content type, content length, connection status, date and time etc. There are various formats in which the logs can be generated and to think from a security point of view then the format should be in a way which can be analysed for detecting errors or some malicious behaviour. There are various sources of logs and are in different formats. Some of the log source includes:

- *Firewall logs:* includes information about the inbound and outbound packets, information about particular srevers e.g. web servers, probing the system, etc.
- *IDS/IPS logs:* provides information about the suspicious packets, host and network based attack statistics, helps in generating attack signatures, etc.
- *Application Server logs:* includes logs generated by web server, mail server (gives connection status, protocol status, etc.), FTP server (gives current logins, file upload or download, etc.), database server (gives information about user activity related to objects accessed, creation of new tables, databases, etc.).

There are various tools that can be used for the generation of network logs such as Wireshark, Capsa, etc. that are being used by organisations for tracing and logging the network activities.

To make the best use of all the logs it need to be managed properly and why it needs to be managed is discussed in the following section.

III. Related Work

Many researchers and developers have debated and demonstrated many approaches which may help in understanding threats and their detection along with prevention. The following are few research works being conducted to improvise the intrusion detection system.

A. An efficient Intrusion Detection Model Based on Fast Inductive Learning

In this paper the data set considered is of network which consist of mix data's both normal and attack based and here Inductive learning has considered as the main learning technique and one of the efficient method to analyse future data for novel attacks. In the fast inductive learning method three steps are required: data partition, the growing rule (GrowRule) and rule pruning (PruneRule). The major advantage of using inductive learning over RIPPER algorithm is that former technique modifies time consuming step of growing rule and pruning rule.

They are using two data structures named as feature list and status list shown in figure 2.

The feature list consists of two columns where first column holds all the feature values and the second column holds number of row from which the value came. Status list consist of original grow set still uncovered. In the beginning all items are set TRUE and once the best split applied on all feature then status list is updated setting FALSE all the rows which were not covered by the new split condition. They have prepared an intrusion detection model based on inductive learning shown in figure 3, that consist of Training stage and Detecting stage and along with it they are building the detection model with double profiles.

DOI: 01.IJNS.4.1. 18

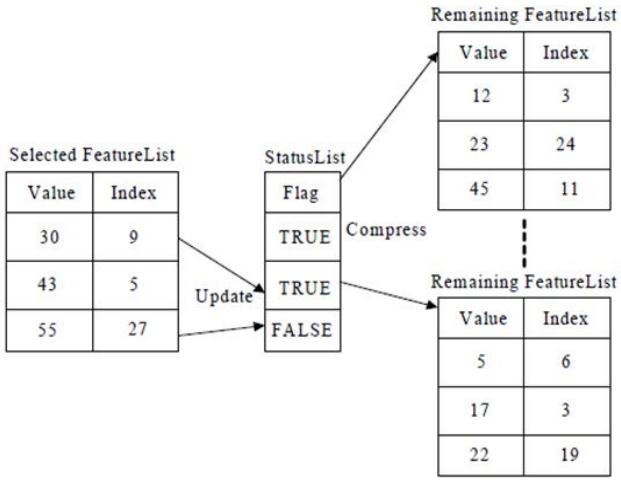

Figure 2. Data Structure in FILMID

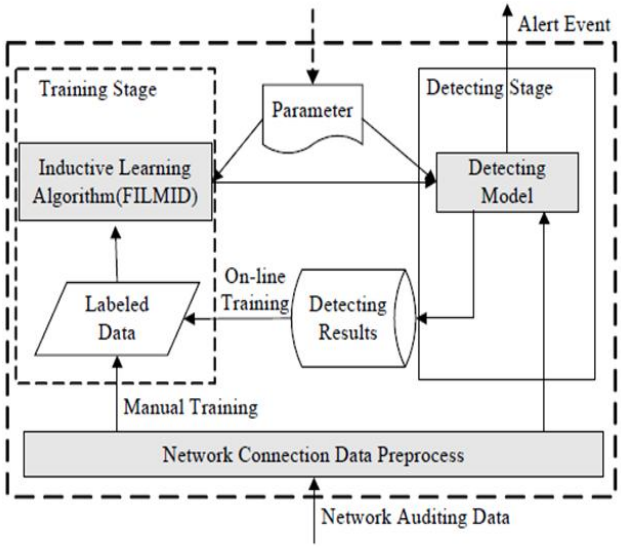

Figure 3. Intrusion Detection model based on Inductive Learning

B. Research of Intrusion Detection System based on Machine Learning

As an evolving technology, intrusion detection technology helps the system to handle the attacks which expand the system administrator's security management capabilities and improved the information security infrastructure integrity [7]. The authors have discussed the emergence of intrusion model being developed since 1986 till present which are based on machine learning. The main focus is on the misuse detection system which has poor capability for unknown attacks.

Basic method to apply machine learning to the misuse detection is to train the learning machine to each known attack of detection object by using attacks as example for generating samples and after the completion of training, learning machine would establish feature contour of the known attack behaviour. Since the machine has good inductive and reasoning ability, it would not only detect known attacks but also variety of attacks and unknown attacks.

The proposed system used machine learning methods to detect the data packets captured from the network and consist of four modules: (i) Network Packet capture module, (ii) Data Pre-processing module, (iii) Misuse rule processing module and (iv) Machine learning modules. The model is shown in the figure 4.

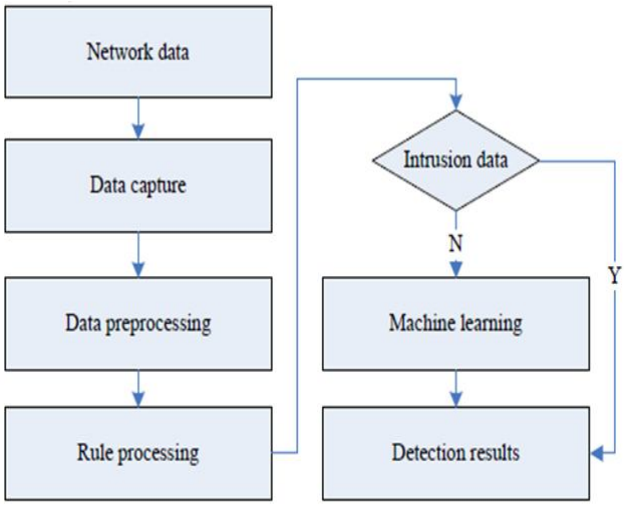

Figure 4. Intrusion detection system model based on Machine Learning

C. Web server Logs Pre-processing for Web Intrusion Detection

The authors of this paper have converged upon analysis of web server log files to detect web attacks. Whenever a user accesses a website, these text files are generated automatically recording information about each user request. The information in the log files are recorded based on a particular log file format. The analysis of such files can reveal patterns of web attacks. They have defined four types of server logs:

i. Access log file: contain information about incoming request and get information about client of the server
ii. Error log file: contain internal server errors and help the administrator to correct the content on site or detect the malicious activity
iii. Agent log file: contain information about users' browser, operating system and browser version
iv. Referrer log file: contain information about the link that redirects the user to a particular site.

The authors have defined the Log file format into three parts: NCSA format that records the basic information about the user request with time field recorded as local time, W3C extended format that is customized and can contain more information than NCSA format having time recorded in GMT format and IIS format which is fixed and fields are separated by commas that make them easy to read. Then these logs go through a log file pre-processing, where they eliminate certain data that is unnecessary and could affect detection of web attacks. Then they apply mining algorithm to detect attacks efficiently.

Figure 5 is showing the steps of logs file pre-processing by which they achieve web intrusion detection. In the initial step which is novel they have combined two different format files into one using XML and achieved with the help of two algorithms.

DOI: 01.IJNS.4.1.18

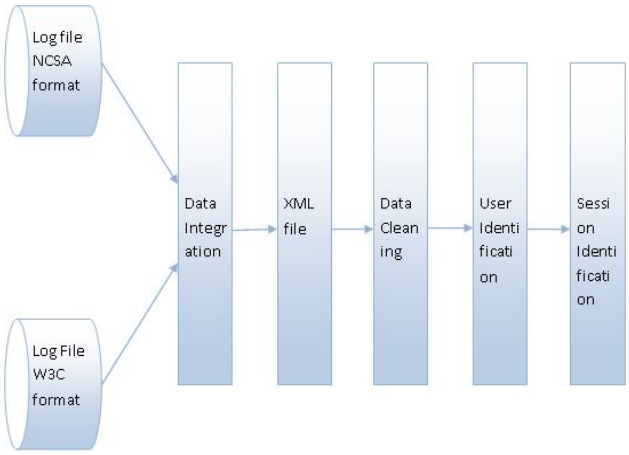

Figure 5. Logs file Pre-processing

D. Hybrid Network Intrusion Detection System using Expert Rule based approach

A.S. Aneetha et al. have proposed a framework which detects the ability to detect intrusion in real time environment. The misuse detection system is being used for known attacks while unknown attacks are detected through anomaly based detection. In the paper the detection rate of the hybrid system has been found to increase as the unknown attack percentage increases whereas in misuse, detection rate is found to decrease and in anomaly detection rate remains constant [9]. The proposed hybrid detection system is designed using rule base for misuse detection and clustering based analysis for anomaly detection. The framework has been divided into different modules such as data capturing, misuse detection, Rule Base, data preprocessing, clustering techniques and expert system. The architecture of the system is defined below with the steps:

- The data is captured from its real time environment using Wireshark tool from the Mac layer.
- The captured data is then compared to the rule base for identifying pre defined pattern of attacks.
- Then comes data pre processing phase where data is classified as attacks previously not considered
- Following the data preprocessing step cluster based technique is applied on large set of data's having low false alarm rate. Grouping is done into different clusters
- After cluster formation, the data was classified as intrusion is given to this module. After the decision of expert system the rules are coded as a new attack and given to rule base for updating.

E. Sharing Computer Network logs for Security and Privacy: A Motivation for New Methodologies of Anonymization

The authors of the paper have discussed about current trends of sharing logs to study them and gather all the information so that they can understand the sensitivity of the information contained within the logs. They have come up with the action items of National Strategy to Secure Cyberspace (NSSC) explicitly listing sharing as the highest priority i.e. data sharing within the government, within industry sectors and between the government and industry. Three of them are directly focusing over the log data sharing: 2: "Provide for the development of tactical strategic analysis of cyber attacks and vulnerability assessments"; item 3: "Encourage the development of a private sector capability to share a synoptic view of the health of cyberspace", and item 8: "Improve and enhance public/private information sharing involving cyber-attacks, threats, and vulnerabilities". They have made client server architecture for the framework: a client, who can be a system administrator or a regular user requests data from a server which is called the Coordinator. The architecture can be depicted in the following figure 7.

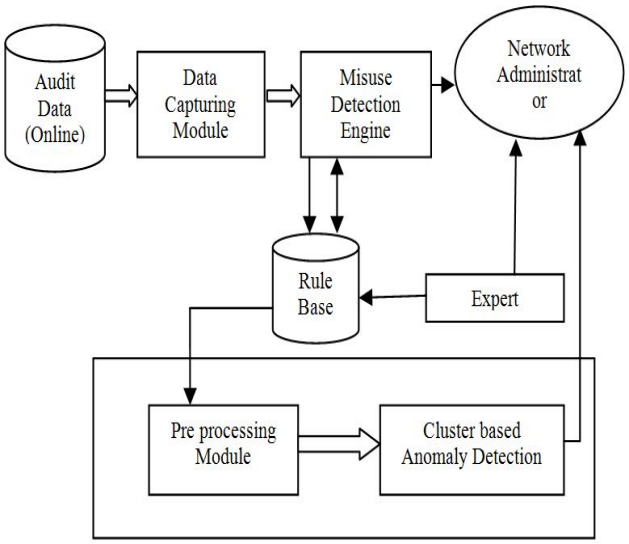

Figure 6. System Architecture

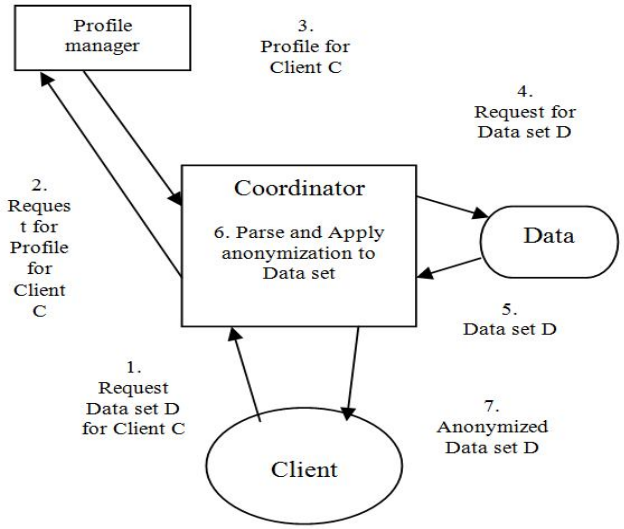

Figure7. System Architecture and how information is passed between components upon requests for specific logs

As shown in figure to answer a request, the coordinator will have to communicate with two other entities: the Profile manager and the Data Store. The profile manager stores profiles that determine the anonymization level for a client. In order to apply the profile to a dataset, the coordinator determines what fields are in the data and then actually apply anonymization algorithm to a particular field [7].

DOI: 01.IJNS.4.1.18

IV. Need of Log Management

With the rapidly increasing use of computer resources and the network the need for improving its security is also essential in best possible manner. The system and network administrators monitoring the system resources and network activities may oversight some events occurring in the system and the network. Thus, in order to minimise the occurrence of losses logs can be generated and analysed with proper management. Since logs keep a track of all the activities occurring in a system there is a rare chance of missing an activity that may cause intrusion in the system or the network. For effective utilisation of logs, they need to be properly managed. Log management helps in storage, analysis and monitoring. It can benefit an organisation in several different ways:

- The effective log management provides the information required for threat detection.
- It helps an organisation to perform automated and cost effective audits.
- Log management helps in keeping a trail of the intruders and the sensitive information that may have been accessed by them.
- It helps in monitoring the network and thereby preventing unauthorised access.
- With regular log management maintenance cost of the system is reduced.
- Logs can be used for forensic analysis by an organisation, so their proper management is required.
- Logs can be used for measuring the performance of an application.
- It helps in archival of huge amount of data produced daily for backup without redundancy

With the large amount of data produced daily in an organisation, effective management of the data is required for securing the sensitive facts and information of an organisation. Thus this information must be effectively managed by preparing logs of data. This will help in providing security to a system in a best possible manner.

V. Architecture that Supports Log Management

For understanding a more effective utilisation of logs here we present an architecture specific to logs for securing network resources. These logs can be analysed for formulation of certain security policies for identification of various attacks.

A. Network Architecture

Figure 8 on the other hand is focusing on the network logs, how we can manage logs generated over network, what all fields can be captured in the network. For achieving all the questions being raised, we first make a website whose backup is kept in the database as soon it is developed and the moment the client continues to move further its log is generated in a separate file which can be further used for analysis. The header field used is IP Address which provides the current IP address of the particular user which logged on into the system. The moment it is observed that a user is accessing the page up to a limited number of times which might affect performance of the service provided by that particular website.

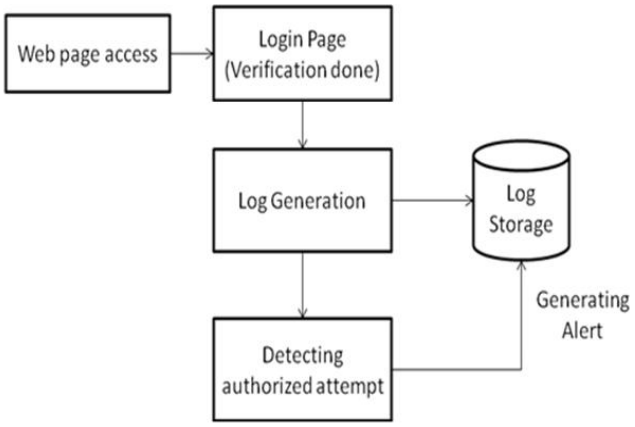

Figure 8.Flow of network log analysis

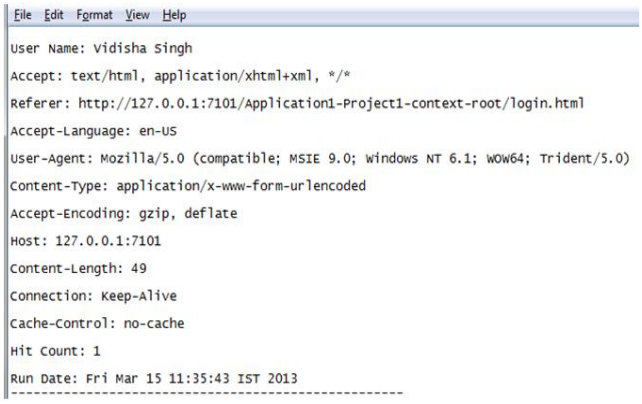

Figure 9. Network log

The network logs comprises of following header information: username, accept, referrer, accept-language, user-agent, content type, accept-encoding, host, content-length, connection, cache-control, hit count and run date. The figure 6 shows the log information being logged as soon as organization site is visited. It is specifically defending denial of service attack which is quite common and can be easily prevented.

The same way log management is highly recommended for networks because any external source can enter the host and thus it requires log analysis by which we can avoid several network attacks. To know more about the log details then here we present the definition of each field:

User Name: Vidisha Singh

Accept: The MIME (Multipurpose Internet Mail Extension) type the browser prefers: Text/html, application/xhtml+xml, */*: type of format it

Referer: the URL of the page containing the link the user followed to get to the current page: http://127.0.0.1:7101/Application1-Project1-context-root/login.html

Accept-Language: the language the browser is expecting, in case the server has versions in more than one language: en-US

User-agent: type of browser, useful if servlet is returning

DOI: 01.IJNS.4.1. 18

browser-specific content: Mozilla/5.0 (compatible; MSIE 9.0; Windows NT 6.1; WOW64; Trident/5.0)
Content-Type: application/x-www-form-urlencoded
Accept-Encoding: the types of data encodings (such as gzip) the browser knows how to decode. Servlets can explicitly check for gzip support and return gzipped HTML pages to browsers that support them, setting the content-encoding response header to indicate that they are gzipped. In many cases, this can reduce page download times by a factor of five or ten: gzip, deflate.
Host: Host and port as listed in the original URL: 127.0.0.1:7101.
Content-Length: for POST messages, how much data is attached: 49 (the message length being sent)
Connection: Use persistent connection? If a servlet gets a Keep-Alive value here, or gets a request line indicating HTTP1.1 (where persistent connections are the default), it may be able to take advantage of persistent connections, saving significant time for Web pages that include several small pieces (images or applet classes). To do this, it needs to send a Content-length header in the response, which is most easily accomplished by writing into a Byte Array Output Stream, then looking up the size just before writing it out: Keep-Alive
Cache-Control: An abstraction for the value of a HTTP Cache-Control response header: no-cache
Hit-Count: The number of times the site has been hit by the user
Run-Date: the current date and time: Fri Mar 15 11:35:43 IST 2013
The following algorithm can be applied for preventing the access of a web page from Denial of Service attack on the basis of header" hit count". In the similar manner various other header fields can be utilised for detection and prevention of other attacks. The algorithm is as follows:
Algorithm: Detection and Prevention of Denial of Service attack and, Network logs generation
Input: Web page access
Output: Network logs + preventing DoS attack

```
Begin
set counter->0
build a hashmap object 'hitCountMap'
get host address and store it as
'ipadd'
if( ipadd is contained in
hitCountMap)
do
hitcount= hitCountMap.get(ipadd)
increase counter linearly
hitCountMap.put(ipadd, counter)
else
hitCountMap.put(ipadd,1)
if(hitCountMap.get(ipadd)<11)
do
create a network log file
store all the information with
respect to header information over
the network
else
block the corresponding ip address
end
```

The main thing to observe through these header files is that they are capable of detecting different types of Attacks that can harm the system adversely. The first header information which can prevent a DoS (Denial of Service) attack is IP header and the hit count. If we are able to count the number of hit rates from a particular host which is requesting that site then we can judge how many requests are coming from a particular IP address. Another measuring unit can be the run-date, by which we can calculate the particular duration within which an attacker may request the page again.

So, for securing the web page we need to generate the log files automatically and put them on surveillance.

IV. Future Scope

There are huge future scopes of network log files as they provide with large amount of information about the network. This information can further be analyzed using various biologically inspired algorithms such as artificial immune system technique, genetics, neural networks algorithms, machine learning, etc. As these biologically inspired approaches are versatile and robust, thus application of these algorithms to the logs can produce an effective log-based system for enhancing the security of system resources and network resources.

V. Conclusions

With the growth in usage of computer resources and network huge amount of data is produced on a daily basis. This data can be used for analysis that would help in identifying the various attacks. Logs generation is the most effective way for analyzing such large amount of data as it keeps a track of all the activities occurring in a system and the network. Log management and analysis is becoming important part for the researches and an effective way for protecting the system from various known and unknown attacks. By analyzing and exploring the information obtained from the headers of the network logs a large number of malicious activities and attacks can be traced down, thereby, protecting the system in every possible manner.

References

[1] Dr. Kees Leune, Logging and Monitoring to Detect Network Intrusions and Compliance Violations, SANS Institute, July 2012.
[2] "Modern Network Security: the migration to Deep Packet Inspection", whitepaper by esoft. Technologies Limited.
[3] Wu Yang, Wei Wan, Lin Guo, Le-Jun Zhang, An efficient Intrusion Detection Model Based on Fast Inductive Learning, Proceedings of sixth international conference on machine learning and cybernetics, August 2007.
[4] Ma Jun, Feng Shuqian, Research of Intrusion Detection System

DOI: 01.IJNS.4.1.18

based on Machine Learning, 2nd International conference on Computer Engineering and Technology, 2010.

[5] Shaimaa Ezzat Salama, Mohamed I. Marie, Laila M. EI-Fangary, Yehia K. Helmy, "Web server Logs Preprocessing for Web Intrusion Detection", published by Canadian Center of Science and Education, July 2011.

[6] A.S. Aneetha, T.S. Indhu, Dr. S.Bose, "Hybrid Network Intrusion Detection System using Expert Rule based approach", CCSEIT-ACM 2012.

[7] Adam Slagell, William Yurcik, Sharing Computer Network logs for Security and Privacy: A Motivation for NewMethodologies of Anonymization, National Center for Supercomputing Applications (NCSA).

[8] Karen Kent, Murugiah souppaya, Guide to Computer Security Log management, recommendations of NIST, September 2006.

[9] Header Information available via http://www.apl.jhu.edu/~hall/java/Servlet-Tutorial/Servlet-Tutorial-Request-Headers.html

DOI: 01.IJNS.4.1. 18